Analyzing Patient No-Show Behavior in Medical Appointments

by Chanatip Srinaul

Table of Contents

01 Introduction

02 Dataset Preview

03 Data Cleaning & Visualization

04 Key Insights & Suggestions

Introduction

Problem

โรงพยาบาลแห่งหนึ่งในประเทศบราซิลมีการนัดผู้ป่วยเพื่อมาพบแพทย์ แต่ผู้ป่วยบางคนที่ไม่มาตามนัดหมาย

Objective

ตรวจสอบและวิเคราะห์ข้อมูลที่เกี่ยวข้องกับผู้ป่วยที่ไม่ตามนัดหมายเพื่อหา insights ที่จะช่วยให้สามารถวางแผนเพื่อลดอัตราการไม่มาตามนัดหมายในอนาคตและปรับปรุงประสิทธิภาพของระบบการแต่งตั้งการดูแลสุขภาพ

Dataset

ข้อมูลของการนัดหมายผู้ป่วย 110,527 ราย 14 variables โดยมี target variable คือ สถานะการมาตามนัดหมายของผู้ป่วย

Dataset Preview

Column ใน Dataset

01. PatientId --> Id ของคนไข้
02. AppointmentID --> Id ของการนัดพบแพทย์
03. Gender --> เพศของผู้ป่วย (ชาย, หญิง)
04. ScheduledDay --> วันที่ทำหนดการนัดพบแพทย์
05. AppointmentDay --> วันที่ผู้ป่วยต้องมาพบแพทย์
06. Age --> อายุของคนไข้
07. Neighbourhood --> ย่านที่อยู่อาศัยของผู้ป่วย
08. Scholarship --> ได้รับทุนจากรัฐหรือไม่
09. Hypertension --> เป็นโรคความดันหรือไม่
10. Diabetes --> เป็นโรคเบาหวานหรือไม่
11. Alcoholism --> ติดสุราหรือไม่
12. Handicap --> พิการหรือไม่
13. SMS_received --> ได้รับ SMS แจ้งเตือนหรือไม่
14. No_show --> สถานะการนัดหมายของผู้ป่วย

ตัวอย่างของ Dataset

PatientId	AppointmentID	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood
2.987250e+13	5642903	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z	62	JARDIM DA PENHA
5.589978e+14	5642503	M	2016-04-29T16:08:27Z	2016-04-29T00:00:00Z	56	JARDIM DA PENHA
4.262962e+12	5642549	F	2016-04-29T16:19:04Z	2016-04-29T00:00:00Z	62	MATA DA PRAIA
8.679512e+11	5642828	F	2016-04-29T17:29:31Z	2016-04-29T00:00:00Z	8	PONTAL DE CAMBURI
8.841186e+12	5642494	F	2016-04-29T16:07:23Z	2016-04-29T00:00:00Z	56	JARDIM DA PENHA

Data Cleaning & Visualization

ตรวจสอบข้อมูลเบื้องต้น

ตรวจสอบภาพรวมของข้อมูลเบื้องต้น พบว่าไม่มีข้อมูลหายไป แต่มีค่าผิดปกติอยู่ในบาง column

ทำความสะอาดข้อมูล

- สร้าง copy ของ dataset ไว้ก่อนจะทำความสะอาดเพื่อไม่ให้เกิดการเปลี่ยนแปลงใน raw data
- ลบ columns ที่ไม่ได้ใช้งาน
- ลบข้อมูลที่มีค่าผิดปกติออก

สร้าง feature ใหม่เพิ่มเติม

- สร้าง Column “WaitingDay” ซึ่งเป็นข้อมูลจำนวนวันที่ผู้ป่วยต้องรอระหว่างวันกำหนดนัดหมายกับวันนัดหมายจริง
- สร้าง Column “Weekday” แทนวันในสัปดาห์ที่ผู้ป่วยต้องมานัดพบแพทย์

สร้าง Visualize และวิเคราะห์ข้อมูล

ทำ EDA โดยการสร้างกราฟและวิเคราะห์ข้อมูลเพื่อหา Insight

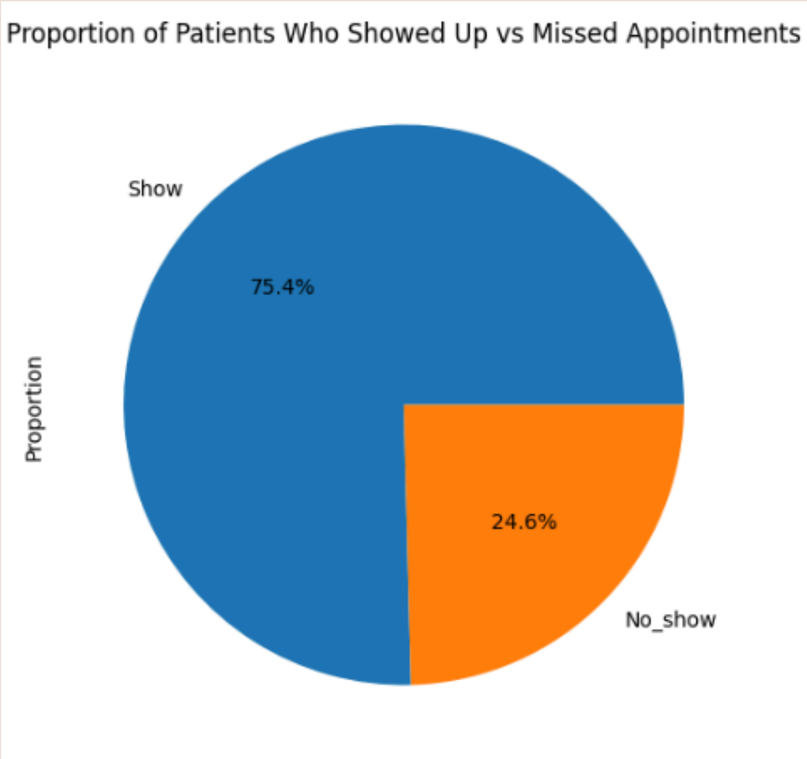
Data Cleaning

ตัวอย่างของ Dataset หลังจากทำความสะอาดแล้ว

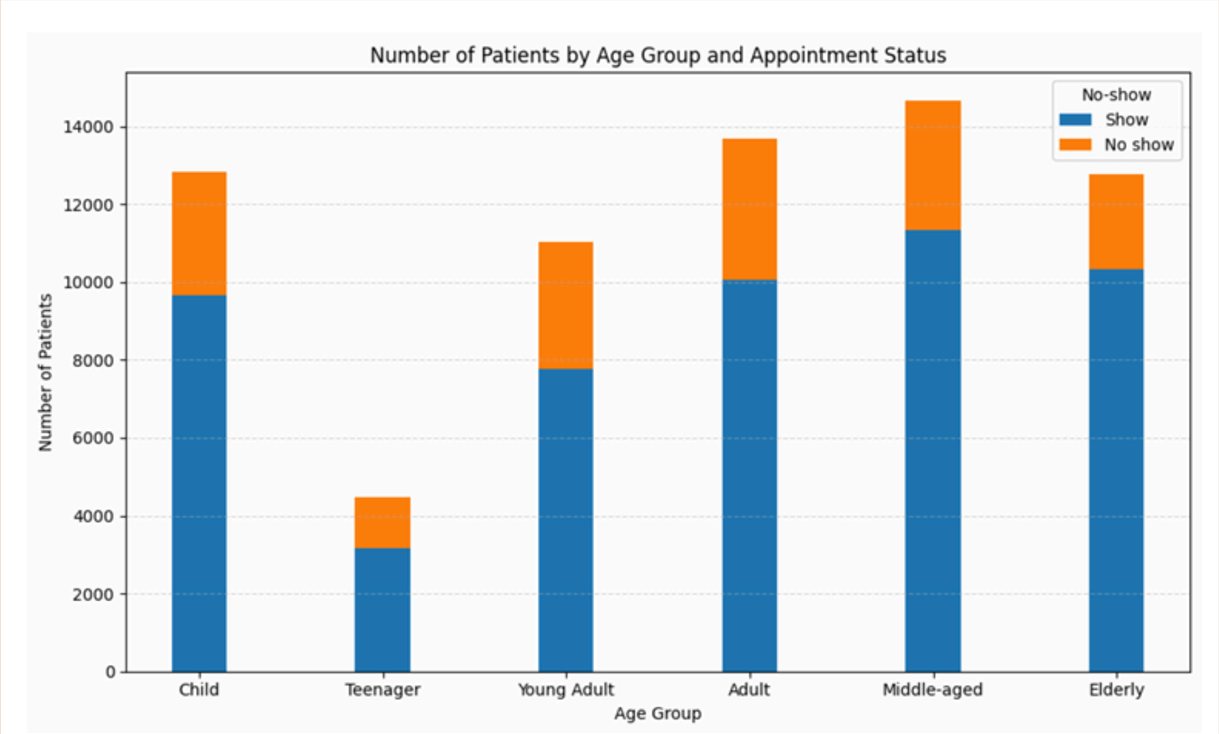
	Gender	Age	Neighbourhood	Scholarship	Hypertension	Diabetes	Alcoholism	Handicap	SMS_received	No_show	WaitingDay	Weekday
0	F	62	JARDIM DA PENHA	0	1	0	0	0	0	0	0	Friday
1	M	56	JARDIM DA PENHA	0	0	0	0	0	0	0	0	Friday
2	F	62	MATA DA PRAIA	0	0	0	0	0	0	0	0	Friday
3	F	8	PONTAL DE CAMBURI	0	0	0	0	0	0	0	0	Friday
4	F	56	JARDIM DA PENHA	0	1	1	0	0	0	0	0	Friday

Data Visaulization

ตัวอย่างของ Visualize ที่สร้างขึ้น



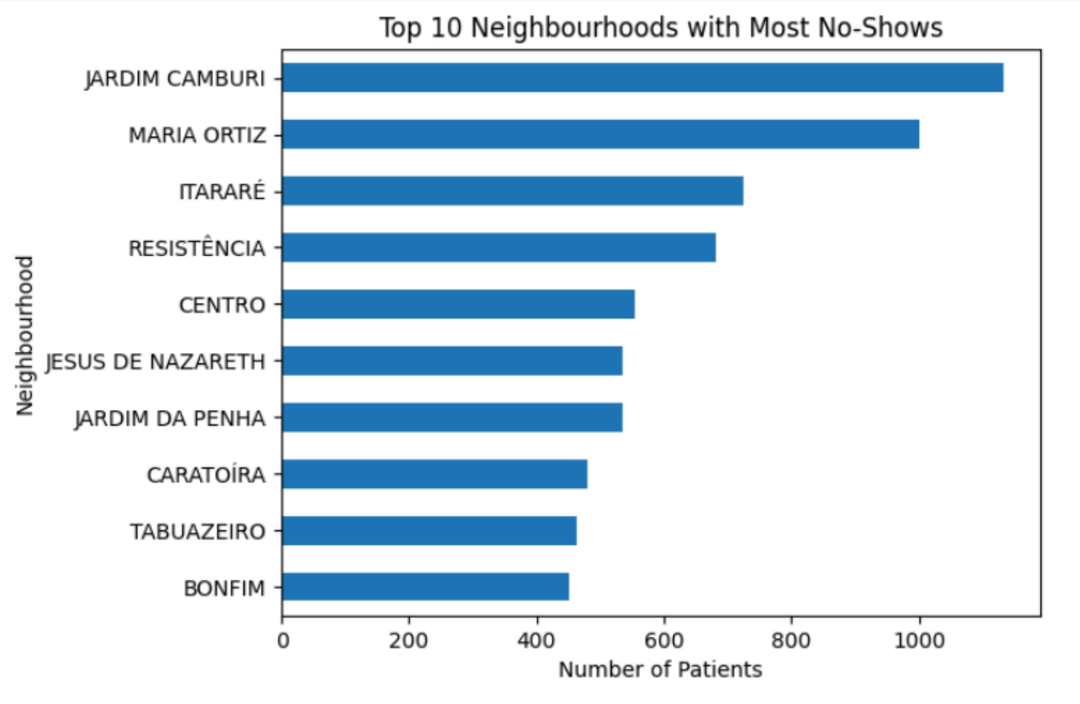
(กราฟแสดงสัดส่วนของผู้ป่วยที่มาตามนัดและไม่มาตามนัด)



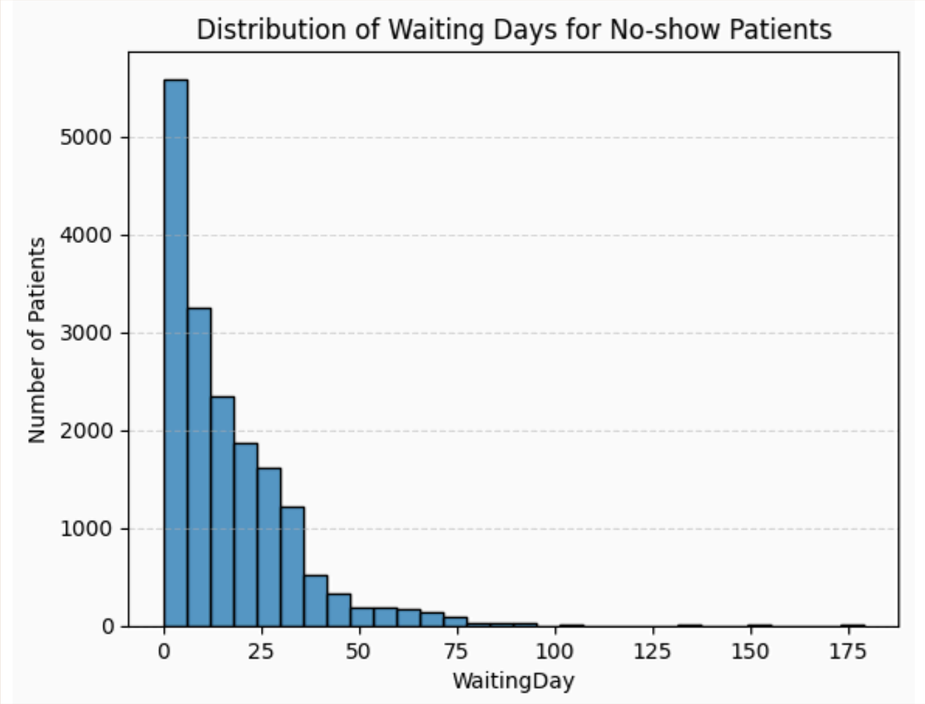
(กราฟแท่งแสดงจำนวนผู้ป่วยที่ไม่มาตามนัดโดยแบ่งตามช่วงอายุ)

Data Visaulization

ตัวอย่างของ Visualize ที่สร้างขึ้น



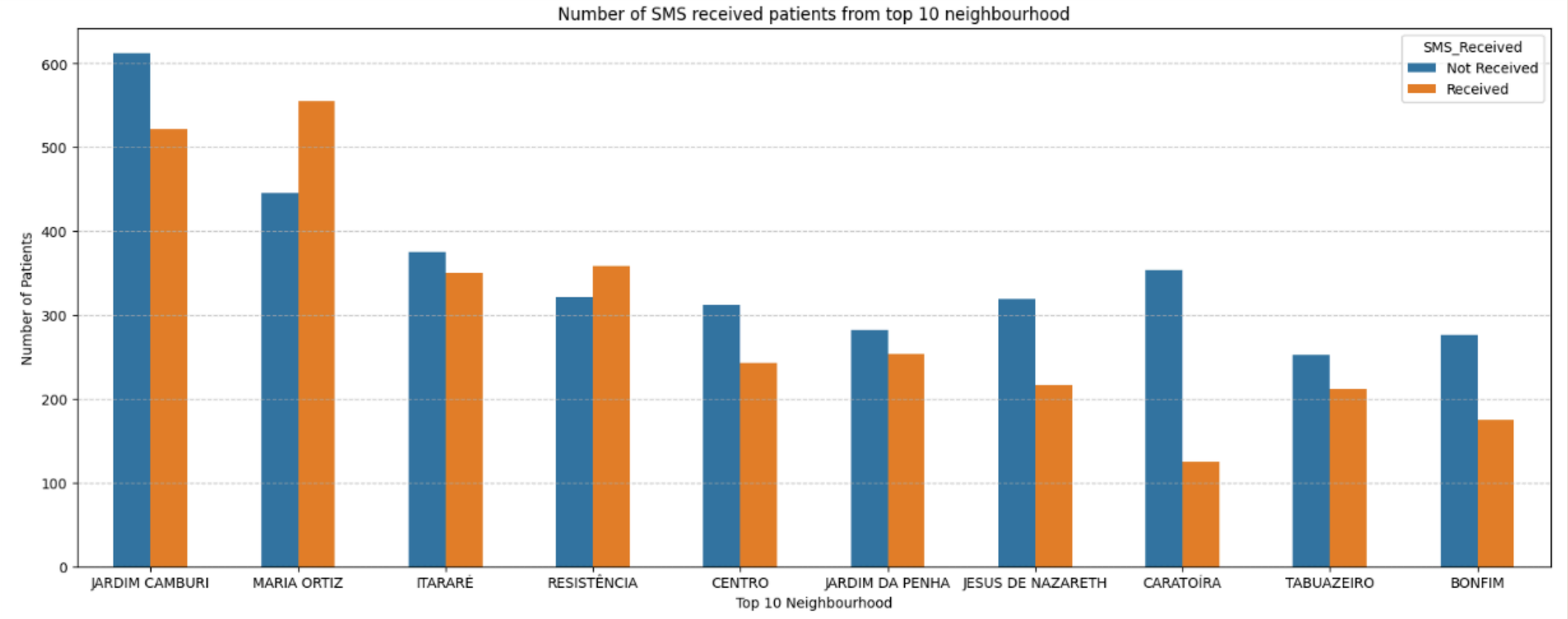
(กราฟแสดงย่านที่อยู่ของผู้ป่วยที่ไม่มาตามนัดมากที่สุด 10 อันดับ)



(กราฟแสดงจำนวนวันที่ผู้ป่วยที่ไม่มาตามนัดต้องรอก่อนจะนัดพบจริง)

Data Visaulization

ตัวอย่างของ Visualize ที่สร้างขึ้น



(กราฟแสดงจำนวนการได้รับ SMS แจ้งเตือนของกลุ่มผู้ป่วยในย่านที่อยู่ที่มีจำนวนผู้ป่วยไม่มาตามนัดสูงสุด 10 ย่าน)

Key Insights & Suggestions

Key Insights

สถานะการนัดหมาย: มีผู้ป่วยที่มาตามนัด 75.4% และมีผู้ป่วยที่ไม่มาตามนัด 24.6%

เพศ: จำนวนของผู้ป่วยหญิงที่ไม่มาตามนัดมากกว่าผู้ป่วยชาย แต่เมื่อดูที่อัตราส่วนของการไม่มาตามนัดในแต่ละเพศ พบว่าทั้งผู้ป่วยชายและหญิงมีอัตราส่วนไม่มาตามนัดเท่าๆกันที่ประมาณ 24%

ช่วงอายุ: กลุ่มผู้ป่วยที่มีจำนวนการไม่มาตามนัดมากที่สุดคือกลุ่มวัยกลางคน (อายุ 45-60 ปี) แต่เมื่อดูที่อัตราการไม่มาตามนัดพบว่ากลุ่มผู้ป่วยที่เป็นผู้ใหญ่ (อายุ 30-45 ปี) มีอัตราการไม่มาตามนัดสูงที่สุดที่ 21.3% ตามด้วยกลุ่มวัยกลางคนที่ 19.36%

ภาวะทางสุขภาพต่างๆ: พบว่าในกลุ่มผู้ป่วยที่ไม่มาตามนัด มีกลุ่มผู้ป่วยที่มีภาวะทางสุขภาพอย่างน้อยหนึ่งอย่างคิดเป็น 21.08%

ระยะเวลาที่ต้องรอ: การนัดหมายที่ผู้ป่วยไม่มาตามนัดส่วนใหญ่เกิดขึ้นในกลุ่มผู้ป่วยที่ต้องรอ 2 วันระหว่างวันที่นัดหมายและวันนัดพบจริง

วันนัดหมายจริง: วันอังคารเป็นวันที่มีผู้ป่วยไม่มาตามนัดมากที่สุด

การแจ้งเตือนทาง SMS: มีผู้ป่วย 33.56% จากผู้ป่วยทั้งหมดที่ได้รับ SMS แจ้งเตือน โดยในกลุ่มผู้ป่วยที่ได้รับ SMS แจ้งเตือนนี้มีผู้ป่วยไม่มาตามนัดสูงถึง 33.21% ซึ่งสูงมากกว่ากลุ่มผู้ป่วยที่ไม่ได้รับ SMS แจ้งเตือนที่มีผู้ป่วยไม่มาตามนัดคิดเป็น 20.2% ของจำนวนผู้ป่วยที่ไม่ได้รับ SMS แจ้งเตือนทั้งหมด

ย่านที่อยู่อาศัย: ย่านที่อยู่อาศัยที่มีผู้ป่วยที่ไม่มาตามนัดมากที่สุดคือ “Jardim Camburi” ซึ่งคิดเป็น 6.42% จากผู้ป่วยที่ไม่มาตามนัดทั้งหมดจาก 81 ย่าน

10 อันดับย่านที่พักอาศัยที่มีผู้ป่วยไม่มาตามนัดมากที่สุด: ในย่านเหล่านี้มีผู้ป่วยที่ไม่มาตามนัด 45.9% ที่ได้รับ SMS แจ้งเตือน อีกทั้งในย่าน Maria Ortiz และ Resistência ยังมีจำนวนผู้ป่วยที่ได้รับ SMS แจ้งเตือนแล้วไม่มาตามนัดสูงกว่าจำนวนผู้ป่วยที่ไม่ได้รับ SMS แจ้งเตือนอีกด้วย

Key Insights & Suggestions

Suggestions

- ให้ความสำคัญกับผู้ป่วยในช่วงอายุ 30-45 ปี เนื่องจากเป็นกลุ่มที่มีอัตราการไม่มาตามนัดสูงที่สุด จึงแนะนำให้ปรับปรุงการแจ้งเตือนการนัดหมายให้เหมาะสมกับพฤติกรรมของกลุ่มอายุนี้มากขึ้น
- ควรทำการศึกษาเพิ่มเติมเพื่อระบุปัจจัยอื่นๆที่อาจส่งผลต่อการมาตามนัดแม้ว่าจะไม่พบความสัมพันธ์ที่ชัดเจนระหว่างเพศหรือภาวะสุขภาพกับการไม่มาตามนัด แต่ปัจจัยอื่นๆ เช่น ขนาดครอบครัว ระดับการศึกษา หรือรายได้ อาจมีส่วนเกี่ยวข้องและควรทำการสำรวจเพิ่มเติม
- ติดตามและวิเคราะห์ย่านที่พักอาศัยที่มีผู้ป่วยไม่มาตามนัดสูงอย่างใกล้ชิด เช่น ย่าน Jardim Camburi ควรวิเคราะห์หาสาเหตุที่เป็นไปได้ เช่น ปัญหาการเดินทาง หรือข้อจำกัดทางภูมิศาสตร์ เพื่อใช้ในการสนับสนุนมาตรการแก้ไขที่ตรงจุดมากขึ้น
- ปรับปรุงระบบแจ้งเตือนทาง SMS ให้มีประสิทธิภาพ เช่น ตรวจสอบความถูกต้องของเบอร์ติดต่อ และช่วงเวลา, ความถี่, และเนื้อหาของ SMS แจ้งเตือน เนื่องจากข้อมูลแสดงให้เห็นว่าผู้ป่วยที่ได้รับ SMS ยังคงมีอัตราการไม่มาตามนัดที่สูงกว่าผู้ที่ไม่ได้รับ SMS

Thank you!

ผู้จัดทำ: นายชนาธิป ศรีนวล
Email: chanatip.srin@gmail.com

Tools ที่ใช้ในโปรเจคนี้

- Python 3.12.0 (Pandas, Matplotlib, Seaborn)
- Jupyter Notebook